Chapter 1

Euclidean Space

1.1 Basic Structure

Def 1.1. Euclidean Space

Euclidean Space, denoted as $\mathbb{R}^n$, is the set of all n-tuples $x = (x_1, \ldots, x_n)$ with each $x_i \in \mathbb{R}$. x is called a **point** or a **vector**. Addition is defined for $\mathbb{R}^n$ for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ as

$$x + y = (x_1 + y_1, \ldots, x_n + y_n).$$

Scalar multiplication is defined for $\lambda \in \mathbb{R}$ as

$$\lambda x = (\lambda x_1, \ldots, \lambda x_n).$$

This definition of Euclidean space lends itself to a vector space structure where the underlying field is $\mathbb{R}$.

Theorem 1.1.

$\mathbb{R}^n$ is a vector space over $\mathbb{R}$.

The proof is ommitted as it follows from the fact that $\mathbb{R}$ is a vector space and its properties are preserved under component wise operations. We further endow $\mathbb{R}^n$ with a **scalar product**.

Def 1.2. Euclidean Scalar Product

The **scalar product** of two vectors $x, y \in \mathbb{R}^n$ is

$$x \cdot y = x_1 y_1 + \ldots + x_n y_n.$$

It can be checked that this defines an inner product over $\mathbb{R}^n$, and thus also gives a natural **Euclidean norm** defined simply as $|x| = \|x\| = \sqrt{x \cdot x}$.

Theorem 1.2. Cauchy-Schwarz Inequality

For all $x, y \in \mathbb{R}^n$, $|x \cdot y| \leq |x||y|$.

Proof. If $y = 0$, then the inequality follows trivially. Assume then that $y \neq 0$. Let $t \in \mathbb{R}$ and $z = x + ty$. Note that $z \cdot z = |z|^2 \geq 0$. Therefore

$$\begin{aligned} 0 \leq (x + ty) \cdot (x + ty) &= x \cdot x + 2t(x \cdot y) + t^2 (y \cdot y) \\ &= |x|^2 + 2t(x \cdot y) + t^2 |y|^2 \\ &= |x|^2 + \left(|y|t + \frac{x \cdot y}{|y|} \right)^2 - \frac{(x \cdot y)^2}{|y|^2} \end{aligned}$$

Since t was arbitrary, taking t to be

$$t = -\frac{x \cdot y}{|y|^2}$$

gives

$$0 \leq |x|^2 - \frac{(x \cdot y)^2}{|y|^2} \implies (x \cdot y)^2 \leq |x|^2 |y|^2.$$

Rooting both sides gives the desired result. ◇

Corollary 1.1. Triangle Inequality

For any $x, y \in \mathbb{R}^n$, $|x + y| \leq |x| + |y|$.

Proof. Note that

$$\begin{aligned} |x + y|^2 &= (x + y) \cdot (x + y) \\ &= |x|^2 + 2x \cdot y + |y|^2 \\ &\leq |x|^2 + 2|x||y| + |y|^2 \qquad (\star) \\ &= (|x| + |y|)^2 \end{aligned}$$

where ($\star$) follows from Cauchy Schwarz. Taking the root of both sides gives the desired result. $\diamond$

Def 1.3. Euclidean distance

The **distance** between $x, y \in \mathbb{R}^n$ is denoted as $d(x, y) := |x - y|$.

Theorem 1.3.

$d(x, y)$ defines a metric on $\mathbb{R}^n$ in the sense that for all $x, y, z \in \mathbb{R}^n$

i) $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$

ii) $d(x, y) = d(y, x)$

iii) $d(x, z) \leq d(x, y) + d(y, z)$

Proof. Both (i) and (ii) follow from the properties of a norm on a vector space. For (iii), note that

$$d(x, z) = |x - z| = |(x - y) + (y - z)| \leq |x - y| + |y - z| = d(x, y) + d(y, z)$$

which was to be shown. $\diamond$

Because $d(x, y)$ is a metric, it is called the **Euclidean metric** and $\mathbb{R}^n$ equipped with d is called a **metric space**.

1.2 Topology of $\mathbb{R}^n$

Def 1.4. Open Ball

Let $r > 0$ and $a \in \mathbb{R}^n$. Then the **open ball** centered at a or radius r is the set

$$B_r(a) = \{x \in \mathbb{R}^n | d(x, a) < r\}.$$

Def 1.5. Open Set

A set $G \subseteq \mathbb{R}^n$ is **open** if for every $a \in G$, $\exists r > 0$ such that $B_r(a) \subseteq G$.

Theorem 1.4.

Open balls are open sets.

Proof. Let $b \in B_r(a)$. That is $|b-a| < r$. Take $\rho = r - |a-b| \geq 0$ and consider some $x \in B_\rho(b)$. Then $|x-b| < \rho = r - |a-b|$ and

$$|x-a| \leq |x-b| + |b-a| < r - |a-b| + |b-a| = r.$$

Therefore $x \in B_r(a)$, meaning $B_\rho(b) \subseteq B_r(a)$. Hence $B_r(a)$ is open. ◇

Theorem 1.5. $\mathbb{R}^n$ is a topology

The following hold in $\mathbb{R}^n$

i) Let $(G_\alpha)_{\alpha \in J}$ be a collection of open sets. Then $\bigcup_{\alpha \in J} G_\alpha$ is open.

ii) Let $(G_\alpha)_{\alpha \in J}$ be a *finite* collection of open sets. Then $\bigcap_{\alpha \in J} G_\alpha$ is open.

Proof.

i) Let $x \in \bigcup_{\alpha \in J} G_\alpha$. Then there is some G_α such that $x \in G_\alpha$. This set must be open, thus there is some $r > 0$ such that $B_r(x) \subseteq G_\alpha$. But note that $G_\alpha \subseteq \bigcup_{\alpha \in J} G_\alpha$. Thus the union is open.

ii) If $\bigcap_{\alpha \in J} G_\alpha = \varnothing$, then trivially the intersection is open. Assume then that $x \in \bigcap_{\alpha \in J} G_\alpha \neq \varnothing$. Then $x \in G_\alpha$ for all $\alpha \in J$. Thus there is a collection of radii r_α such that $B_{r_\alpha}(x) \subseteq G_\alpha$. Taking $r = \min_{\alpha \in J} r_\alpha$, the ball $B_r(x) \subseteq B_\alpha(x) \subseteq G_\alpha$ for all $\alpha \in J$. Thus the intersection is open.

◇

Remark. The intersection of an infinite collection of open sets is not neccesarily open. Consider the family of open intervals in $\mathbb{R}$ of the form

$$J_n = \left(-\frac{1}{n}, \frac{1}{n}\right).$$

Note that $\bigcap J_n = \{0\}$ which is not open.

Def 1.6. Neighborhood

Let $a \in \mathbb{R}^n$. A **neighborhood** of a is an open set $G \subseteq \mathbb{R}^n$ such that $a \in G$. Often the term *nbhd* is used as a shorthand.

Remark. If G is a nbhd of a, then $\exists r > 0$ such that $B_r(a) \subseteq G$.

Def 1.7. Interior

The **interior** of a set $A \subseteq \mathbb{R}^n$ is defined as

$$\operatorname{int}(A) := \{x \in \mathbb{R}^n : x \text{ has a nbhd } G \subseteq A\}.$$

Example.

i) $\operatorname{int}([a, b)) = (a, b)$ since any nbhd of a will contain points outside of the interval.

ii) Let $A = \{(x, y) \in \mathbb{R}^2 : x, y \geq 0\}$. Then $\operatorname{int}(A) = \{(x, y) \in \mathbb{R}^2 : x, y > 0\}$ as any point along the axes fail by the same reasoning as above.

iii) $\operatorname{int}(\mathbb{Q}) = \varnothing$ because there will always be an irrational x in any ball based around a rational number.

Theorem 1.6.

For any $A \subseteq \mathbb{R}^n$

i) $\operatorname{int}(A)$ is open

ii) $\operatorname{int}(A)$ is the largest open set contained in A

Proof. Let $x \in \operatorname{int}(A)$. Then there is some nbhd G such that $G \subseteq A$. Let $y \in G$. Since G is open, G is a nbhd of y as well hence $y \in \operatorname{int}(A)$. Therefore $G \subseteq \operatorname{int}(A)$ meaning $\operatorname{int}(A)$ is open. ◇

Def 1.8. Closed set

A set $F \subseteq \mathbb{R}^n$ is **closed** if its complement F^c is open.

Example.

i) Both $\varnothing$ and $\mathbb{R}^n$ are closed

ii) $[a, b]$ is closed for all $a \neq b$

iii) $[a, \infty)$ is closed since $[a, \infty)^c = (-\infty, a)$ which is open

Theorem 1.7.

For every $a \in \mathbb{R}^n$ and $r > 0$, the closed ball $B_r[a] = \{x \in \mathbb{R}^n : |x - a| \leq r\}$ is closed in $\mathbb{R}^n$.

Proof. If $B_r[a]^c = \{x \in \mathbb{R}^n : |x - a| > r\}$ is open, then the desired result is achieved. Let $x \in B_r[a]^c$. Since $|x-a| > r$, then $\exists \rho > 0$ such that $|x-a| = r+\rho$. Take $y \in B_\rho(x)$. Then

$$\begin{aligned} |x-a| \leq |x-y| + |y-a| &\implies |y-a| \geq |x-a| - |x-y| \\ &\implies |y-a| > |x-a| - \rho = r \end{aligned}$$

Therefore $y \in Br[a]^c$, meaning $B_r[a]$ is open. ◇

Def 1.9. Cluster Point

Let $A \subseteq \mathbb{R}^n$. Then $x \in \mathbb{R}^n$ is a **cluster point** of A if every nbhd of x intersects A. Equivalently, x is a cluster point of A iff for every $r > 0$, $B_r(x) \cap A \neq \varnothing$.

Remark. Any point $x \in A$ is a cluster point since $x \in B_r(x)$ for any $r > 0$ and hence $\varnothing \neq \{x\} \subseteq B_r(x) \cap A$. However, it need be that a cluster point is an element of A.

Example.

i) Consider $A_1 = \left\{\frac{1}{n} : n = 1, 2, 3, \ldots\right\} \subseteq \mathbb{R}$. The point 0 is a cluster point since for any $r > 0$, $\exists n \geq 1$ such that $\frac{1}{n} < r$. However $0 \notin A_1$

ii) Consider $A_2 = \left\{(x, y) \in \mathbb{R}^2 : x, y > 0\right\}$. The set of all cluster points is $\left\{(x, y) \in \mathbb{R}^2 : x, y \geq 0\right\}$

Def 1.10. Closure

The set of all cluster points for a set $A \subseteq \mathbb{R}^n$ is the **closure** of A, denoted as $\overline{A}$.

For Example 3, the closure of A_1 is $\overline{A_1} = A_1 \cup \{0\}$ and the closure of A_2 is $\left\{(x, y) \in \mathbb{R}^2 : x, y \geq 0\right\}$. These sets are both closed, a fact which holds in general.

Theorem 1.8. Properties of Closure

Let $A \subseteq \mathbb{R}^n$. Then

i) $\overline{A}^c = \text{int}(A^c)$

ii) $\overline{A}$ is closed

iii) $\overline{A}$ is the smallest closed set containing A

iv) F is closed if and only if $F = \overline{F}$

Proof.

i) Let $x \in \overline{A}^c$. Then x is not a cluster point. Therefore there is some nbhd G of x such that $G \cap A = \varnothing$. Thus $G \subseteq A^c$, hence $x \in \text{int}(A^c)$. Let $x \in \text{int}(A^c)$. Then there is some nbhd H such that $H \subseteq A^c$. Therefore $H \cap A = \varnothing$ meaning x is not a cluster point of A. Thus $x \notin \overline{A} \implies x \in \overline{A}^c$.

ii) From (i), the complement of the closure of a set is the interior of a set. The interior of a set is always open, thus the closure of a set is closed.

iii) Let $F \subseteq \mathbb{R}^n$ such that F is closed and $A \subseteq F$. Note that $A^c \supseteq F^c$ and that F^c is open. Furthermore $\text{int}(A^c)$ is the largest open set contained in A^c, therefore $F^c \subseteq \text{int}(A^c)$. Taking the complement and applying (i) gives $F \supseteq (\text{int}(A^c))^c = \overline{A}$.

iv) Assume that F is closed. Since trivially $F \subseteq F$, by (iii) it follows $\overline{F} \subseteq F$. By definition, $F \subseteq \overline{F}$. Thus $F = \overline{F}$. Assume that $F = \overline{F}$. By (ii), $\overline{F}$ is closed and therefore F is closed.

$\diamond$

Theorem 1.9. Closed Set Families

The following hold in $\mathbb{R}^n$

i) Let $(F_\alpha)_{\alpha \in J}$ be a collection of closed sets. Then $\bigcap_{\alpha \in J} F_\alpha$ is closed

ii) Let $(F_\alpha)_{\alpha \in J}$ be a finite collection of closed sets. Then $\bigcup_{\alpha \in J} F_\alpha$ is closed

Proof.

i) Note that by De'Morgan's,

$$\left(\bigcap_{\alpha\in J} F_\alpha\right)^c = \bigcup_{\alpha\in J} F_\alpha^c.$$

Since every F_α^c is open, and the union of a family of open sets is open, then the complement of the intersection is open. Hence the intersection is closed.

ii) The same application of De'Morgan's gives the desired result.

$\diamond$

Remark. Consider the family of closed sets $F_n = \left[-1+\frac{1}{n}, 1-\frac{1}{n}\right]$. Note that

$$\bigcup_{n=1}^{\infty} F_n = (-1, 1)$$

hence the infinite union of closed sets is not necessarily closed.

Def 1.11. Boundary

For a set $A \subseteq \mathbb{R}^n$, the **boundary** of A is $\partial A := \overline{A} \cap \overline{A^c}$. Equivalently, the boundary is $\partial A = \overline{A} \setminus \operatorname{int}(A)$.

Example. For an open ball $B_r(a) \subseteq \mathbb{R}^n$, its boundary is

$$\begin{aligned}\partial B_r(a) &= \overline{B_r(a)} \setminus \operatorname{int}(B_r(a)) \\ &= \{x \in \mathbb{R}^n : |x-a| \le r\} \setminus \{x \in \mathbb{R}^n : |x-a| < r\} \\ &= \{x \in \mathbb{R}^n : |x-a| = r\}\end{aligned}$$

Chapter 2

Sequences

Def 2.1. Sequence

A **sequence** in $\mathbb{R}^n$ is a map $f : \mathbb{N} \to \mathbb{R}^n$ where

$$f(k) := x^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)}) \in \mathbb{R}^n$$

and is denoted by $\left\{x^{(k)}\right\}$ or $\left(x^{(k)}\right)_{k=1}^{\infty}$.

Def 2.2. Convergence

Let $\left\{x^{(k)}\right\}$ be a sequence in $\mathbb{R}^n$. Then $x^{(k)}$ **converges** to a point $a \in \mathbb{R}^n$ if

$$\lim_{k\to\infty} \left|x^{(k)} - a\right| = 0.$$

Equivalently

1. For all $\varepsilon > 0$, $\exists K \in \mathbb{N}$ such that for all $k \geq K$,

$$|x^{(k)} - a| \leq \varepsilon$$

2. For every nbhd V of a, there is some K such that $x^{(k)} \in V$ for all $k \geq K$

Theorem 2.1.

A sequence $\left\{x^{(k)}\right\}$ converges to $a \in \mathbb{R}^n$ iff for every $1 \leq j \leq n$, the sequence $\left\{x_j^{(k)}\right\}$ converges to a_j.

Proof. For $y \in \mathbb{R}^n$, note that

$$|y_j| \leq \|y\| \leq \sum_{i=1}^{n} |y_i|.$$

Therefore

$$0 \leq \left|x_j^{(k)} - a_j\right| \leq \left\|x^{(k)} - a\right\| \leq \sum_{j=1}^{n} \left|x_j^{(k)} - a_j\right|.$$

Assuming the forward direction, it follows that $\left\|x^{(k)} - a\right\| \to 0$ thus by the squeeze lemma $\left|x_j^{(k)} - a_j\right| \to 0$. Assuming the reverse direction, it follows that $\sum_{j=1}^{n} \left|x_j^{(k)} - a_j\right| \to 0$ which again by squeeze lemme means $\left\|x^{(k)} - a\right\| \to 0$, which was to be shown. ◇

Theorem 2.2. Cluster Point ⇔ Limit Point

Let $A \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Then $x \in \overline{A}$ iff there exists a sequence $\left\{x^{(k)}\right\}$ in A that converges to x.

Proof.

⇐) Suppose such a sequence exists. Then for every nbhd V of x, there is some K such that $x^{(k)} \in V$ for all $k \geq K$. Since $x^{(k)} \in A$ for all k, then it follows $A \cap V \neq \varnothing$, thus $x \in \overline{A}$.

⇒) Suppose $x \in \overline{A}$. Then for any $k \geq 1$, $B_{k^{-1}}(x) \cap A \neq \varnothing$. Therefore for each k, pick some $x^{(k)} \in B_{k^{-1}}(x) \cap A$. Then

$$\left\|x^{(k)} - x\right\| < \frac{1}{k} \to 0$$

thus $\left\{x^{(k)}\right\}$ is such a sequence.

◇

Def 2.3. Bounded Sequence

A sequence $\left\{x^{(k)}\right\}$ in $\mathbb{R}^n$ is **bounded** if there exists $M \geq 0$ such that $\left\|x^{(k)}\right\| \leq M$ for all $k \geq 1$.

Def 2.4. Subsequence

Let $\left(x^{(k)}\right)_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n$ and $\phi : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function. Then the sequence $\left(x^{(\phi(l))}\right)_{l=1}^{\infty}$ is a **subsequence** of the original sequence. The subindex will be denoted simply as $k_l \equiv \phi(l)$.

Theorem 2.3.

For any subsequence, $k_l \geq l$ for all $l \geq 1$.

Proof. Proceed with induction. Note that $k_1 \geq 1$ for any subsequence, thus the base case holds. Assume for some fixed l that $k_l \geq l$. Since the associated ϕ is strictly increasing

$$k_{l+1} > k_l \geq l \implies k_{l+1} \geq l+1$$

which was to be shown. $\diamond$

Theorem 2.4. Bolzano-Weierstrass

Every bounded sequence in $\mathbb{R}^n$ has a convergent subequence.

Proof. Let $\left\{x^{(k)}\right\}$ be a bounded sequence in $\mathbb{R}^n$. Note that $\left|x_j^{(k)}\right| \leq \left\|x^{(k)}\right\|$ for all $1 \leq j \leq n$ and $k \geq 1$. Therefore each element wise sequence is bounded. Thus by Bolzano-Weierstrass in $\mathbb{R}$, the first component has a convergent subsequence with index k_{j_1}. The second component under this index must also be bounded, thus Bolzano-Weiestrass applies to it as well to get another index k_{j_2}. This can be continued until an index k_{j_n} is reached. It is guaranteed by its construction that every component of $x^{(k_{j_n})}$ converges. Thus the subsequence $\left\{x^{k_{j_n}}\right\}$ converges. $\diamond$

Def 2.5. Cauchy Sequence

A sequence $\{x^{(k)}\}$ in $\mathbb{R}^n$ is **Cauchy** if $\forall \varepsilon > 0$, there exists $K \in \mathbb{N}$ such that

$$\left\|x^{(k)} - x^{(l)}\right\| \leq \varepsilon \quad l, k \geq K.$$

Theorem 2.5.

Cauchy sequences in $\mathbb{R}^n$ are bounded.

Proof. Let $\{x^{(k)}\}$ be a Cauchy sequence in $\mathbb{R}^n$. Take $\varepsilon = 1$ and $l = k$. Then as in the definition, for all $k \geq K$,

$$\left\|x^{(k)}\right\| \leq \left\|x^{(k)} - x^{(K)}\right\| + \left\|x^{(K)}\right\| \leq \left\|x^{(K)}\right\| + 1.$$

Take $M = \max\left\{\left\|x^{(1)}\right\|, \ldots, \left\|x^{(K-1)}\right\|, \left\|x^{(K)} + 1\right\|\right\}$. This is well defined since K is finite, and bounds every element. ◇

Theorem 2.6. Completeness

A sequence converges iff it is Cauchy.

Proof. Let $\{x^{(k)}\}$ be a sequence in $\mathbb{R}^n$.

$\Rightarrow$) Assume $\{x^{(k)}\}$ converges with limit $a \in \mathbb{R}^n$. Then for all $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that $\left\|x^{(k)} - a\right\| \leq \frac{\varepsilon}{2}$ for $k \geq 2$. Now consider $k, l \geq K$. Note then that

$$\left\|x^{(k)} - x^{(l)}\right\| \leq \left\|x^{(k)} - a\right\| + \left\|x^{(l)} - a\right\| \leq 2 \cdot \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\{x^{(k)}\}$ is Cauchy.

$\Leftarrow$) Assume $\{x^{(k)}\}$ is Cauchy. Then it is bounded and thus by Bolzano-Weierstrass, it has a convergent subsequence $\{x^{(k_j)}\}$ with some limit $a \in \mathbb{R}^n$. Take $\varepsilon > 0$. Then there exists $J \in \mathbb{N}$ such that for all $j \geq J$

$$\left\|x^{(k_j)} - a\right\| \leq \frac{\varepsilon}{2}.$$

Since $\{x^{(k)}\}$ is Cauchy, there exists $K \in \mathbb{N}$ such that for all $k, l \geq K$

$$\left\|x^{(k)} - x^{(l)}\right\| \leq \frac{\varepsilon}{2}.$$

Take $N = \max\{K, J\}$. Note that $k_j > j$ and so if $k \geq N$ and $j \geq N$

$$\left\|x^{(k)} - a\right\| \leq \left\|x^{(k)} - x^{(k_j)}\right\| + \left\|x^{(x_j)} - a\right\| \leq 2 \cdot \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\left\{x^{(k)}\right\}$ converges.

$\diamond$

Chapter 3

Functions

Def 3.1. Function Terminology

Consider a function $f : D \to \mathbb{R}^p$ where $D \subseteq \mathbb{R}^n$. The **domain** of f is D and the **range** of f is $f(D) := \{f(x) : x \in D\}$.

Example. A function $L : \mathbb{R}^n \to \mathbb{R}$ is a *linear function* if it is of the form

$$L(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

where $c_i \in \mathbb{R}$. These functions are *linear* in their arguments, meaning $L(ax + by) = aL(x) + bL(y)$ for $a, b \in \mathbb{R}$.

Example. A function $Q : \mathbb{R}^n \to \mathbb{R}$ is a *quadratic form* if it is of the form

$$Q(x) = \sum_{i \leq j,k \leq n} c_{jk}x_jx_k$$

where $c_{jk} \in \mathbb{R}$. For example $\|\cdot\|$ is a quadratic form.

Example. A function's domain need not be all of $\mathbb{R}^n$. Consider $f : D \to \mathbb{R}^2$ where $D = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, (x, y) \neq (0, 0)\right\}$ and

$$f(x, y) = \left(\sqrt{4 - x^2 - y^2}, \log\sqrt{x^2 + y^2}\right).$$

This function is well defined on D.

Def 3.2. Limit Point

Let $A \subseteq \mathbb{R}^n$. Then $x \in \mathbb{R}^n$ is a **limit point** of A if for all $r > 0$, $(B_r(x) \setminus \{x\}) \cap A \neq \varnothing$. The set of all limits points of A is denoted as A'.

Remark. If x is a limit point of A, then x is a cluster point. That is, $x \in \overline{A}$. However, the converse is not true.

- Consider $A = B_1(0) \cup P$ for some $P \notin B_1[0]$. Note that $\overline{A} = B_1[0] \cup P$ but $A' = B_1[0]$.
- Consider $A = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. Note that $\overline{A} = A \cup \{0\}$ but $A' = \{0\}$.

Def 3.3. Limit

Let $D \subseteq \mathbb{R}^n$, $f : D \to \mathbb{R}^n$, and $a \in \mathbb{R}^n$ be a limit point of D. Then $f(x)$ has a **limit** to $b \in \mathbb{R}^p$ when x tends to a if for all $\varepsilon > 0$, $\exists \delta > 0$ such that

$$0 < |x - a| \leq \delta \implies |f(x) - b| \leq \varepsilon.$$

This limit is denoted as $\lim\limits_{x \to a} f(x) = b$.

Remark. Limit points must be used when defining limits to ensure that $0 < |x - a| < \delta$ is a non empty set.

Theorem 3.1.

Let $D \subseteq \mathbb{R}^n$, $f : D \to \mathbb{R}^p$ and $a \in \mathbb{R}^n$ be a limt point of D. Then

$$\lim_{x \to a} = b \Leftrightarrow f\left(x^{(k)}\right) \to b, k \to \infty$$

for every $x^{(k)} \to a$ and $x^{(k)} \neq a$.

Corollary 3.1.

Using the same setup as above and letting $f_j : \mathbb{R}^n \to \mathbb{R}$ and b_j be the j^{th} components, then

$$\lim_{x \to a} f(x) = b \Leftrightarrow \lim_{x \to a} f_j\left(x^{(k)}\right) = b_j, 1 \leq j \leq n.$$

Corollary 3.2.

Let $f : D \to \mathbb{R}^p$ and $p : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$ be a limit point of D. If $f(x) \to b$ and $p(x) \to d$ as $x \to a$ then

1. $f(x) + p(x) \to b + d$
2. $\lambda f(x) \to \lambda b$ for all $\lambda \in \mathbb{R}$
3. If $p = 1$, then $f(x)p(x) \to bd$ and $\frac{f(x)}{p(x)} \to \frac{b}{d}$ if $d \neq 0$.

3.1 Continuity

Def 3.4. Continuity

Let $D \subseteq \mathbb{R}^n$ and $a \in D$. A map $f : D \to \mathbb{R}^p$ is **continuous at a** if for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|x - a\| \leq \delta \implies \|f(x) - f(a)\| \leq \varepsilon.$$

If f is continuous at all $x \in D$, then f is **continuous on D**.

Remark. If a is a limit point, then f is continuous iff $\lim_{x \to a} f(x) = f(a)$.

Theorem 3.2. Sequential Continuity

Let $f : D \to \mathbb{R}^p$, $a \in D$. Then f is continuous at a iff for every sequence $x^{(k)}$ in D that converges to a, $\lim_{k \to \infty} f\left(x^{(k)}\right) = f(a)$.

Corollary 3.3.

$f : D \to \mathbb{R}^p$ is continuous at $a \in D$ iff f_j is continuous at a for all $1 \leq j \leq n$.

Corollary 3.4.

If $f, p : D \to \mathbb{R}^p$ are continuous at $a \in D$, then $f + p$, fp are continuous and if $p \neq 0$ then $\frac{f}{p}$ is continuous.

Example. Consider the map $f : \mathbb{R}^2 \to \mathbb{R}$ where

$$f(x) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Consider the continuity of f at $x = 0$. Note that $f(x_1, 0) \to 0$ as $x_1 \to 0$ and $f(x_1, x_1) \to \frac{1}{2}$ as $x_1 \to 0$. Thus $\lim_{x \to 0} f(x)$ does not exist, and thus f is not continuous at 0.

Example. Let $A \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Define the distance x from A as $d(x, A) = \inf_{a \in A} \|x - a\|$. Then $d(\cdot, A) : \mathbb{R}^n \to \mathbb{R}$ is continuous.

Proof. Let $x_0 \in \mathbb{R}^n$. Note for any $a \in A$ that $d(x_0, A) \leq \|x_0 - a\|$. Let $y \in \mathbb{R}^n$. Note then that

$$\begin{aligned} d(x_0, A) &\leq \|x_0 - a\| \\ &\leq \|x_0 - y\| + \|y - a\| \\ &\leq \|x_0 - y\| + d(y, A) \end{aligned}$$

Therefore $d(x_0, A) - d(y, A) \leq \|x_0 - y\|$. By symmetry, x_0 and y can be swapped and so $d(x_0, A) - d(y, A) \leq \|x_0 - y\|$. Thus $|d(x_0, A) - d(y, A)| \leq \|x_0 - y\|$. Let $\varepsilon > 0$ and $\delta = \varepsilon$. Then for $\|x_0 - y\| \leq \delta$, it follows that

$$|d(x_0, A) - d(y, A)| \leq \|x_0 - y\| \leq \delta = \varepsilon.$$

Therefore $d(\cdot, A)$ is continuous. ◇

Def 3.5. Preimage

Let $f : \mathbb{R}^n \to \mathbb{R}^p$ and $G \subseteq \mathbb{R}^p$. The **preimage** of G under f is $f^{-1}(G) = \{x \in \mathbb{R}^n : f(x) \in G\}$.

Theorem 3.3. Topological Continuity

A function $f : \mathbb{R}^n \to \mathbb{R}^p$ is continuous iff $f^{-1}(G)$ is open (closed) in $\mathbb{R}^n$ for every open (closed) set in $\mathbb{R}^p$.

Lemma 3.1.

For any map $f : A \to B$ and set $F \subseteq B$, $f^{-1}(F^c) = f^{-1}(F)^c$.

Proof. Note that

$$\begin{aligned} x \in f^{-1}(F^c) &\Leftrightarrow f(x) \in F^c \\ &\Leftrightarrow f(x) \notin F \\ &\Leftrightarrow x \notin f^{-1}(F) \\ &\Leftrightarrow x \in f^{-1}(F)^c \end{aligned}$$

which was to be shown. ◇

Proof of Theorem 3.3. We consider the open case first. Suppose G is open. Let $a \in f^{-1}(G)$. Then $f(a) \in G$, thus $\exists \varepsilon > 0$ such that $B_\varepsilon(f(a)) \subseteq G$. That is,

$$\|y - f(a)\| < \varepsilon \implies y \in G. \tag{$\star$}$$

Since f is continuous at a, $\exists \delta > 0$ such that $\forall x \in \mathbb{R}^n, \|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon$. Consider $x \in B_\delta(a)$. Then $\|x - a\| < \delta$, thus by continuity of f and $(\star)$, $f(x) \in G$ meaning $x \in f^{-1}(G)$. Thus $B_\delta(a) \subseteq f^{(-1)}(G)$, meaning $f^{-1}(G)$ is open.

Suppose that $f^{-1}(G)$ is open for all open $G \subseteq \mathbb{R}^p$. Take $a \in \mathbb{R}^n$, $\varepsilon > 0$ and let $G = B_\varepsilon(f(a))$. Note that G is open. Suppose $a \in f^{-1}(G)$. Then $\exists \delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(G)$. Therfore $x \in B_\delta(a) \implies x \in f^{-1}(x) \implies f(x) \in G$. Equivalently,

$$\|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon.$$

Thus f is continuous.

To prove the closed case, suppose f is continuous. Take $F \subseteq \mathbb{R}^p$ closed. Then F^c is open in $\mathbb{R}^p$. Thus $f^{-1}(F^c)$ is open. By lemma 3.1, $f^{-1}(F^c) = f^{-1}(F)^c$, thus $f^{-1}(F)^c$ is open. Therefore $f^{-1}(C)$ is closed. The reverse direction follows by a similar argument as above. ◇

Remark. It is not true generally that a continuous map takes open sets to open sets, nor closed set into closed sets. The zero map $f : \mathbb{R} \to \mathbb{R} : x \mapsto 0$ is continuous, but $f((a, b)) = \{0\}$ which is closed. The map $f : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{x^2}{x^2+1}$ is continuous as well, but $f(\mathbb{R}) = [0, 1)$, meaning both an open and closed set is mapped to a set that is neither open or closed.

3.2 Compactness and Uniform Continuity

Def 3.6. Sequential Compactness

A set $K \subseteq \mathbb{R}^n$ is **sequentially compact** if every sequence $(x^{(k)})$ in K has a convergent subsequence that converges to a point in K.

Example. The closed ball $B_r[a] \subseteq \mathbb{R}^n$ is compact. Let $(x^{(k)})$ be a sequence in $B_r[a]$. Note that $\left\|x^{(k)}\right\| \leq \left\|x^{(k)} - a\right\| + \|a\| \leq r + \|a\|$. Thus $(x^{(k)})$ is bounded. Therefore by Bolzano-Weierstrass there exists a subsequence $(x^{(k_j)})$ in K that converges to some point $x \in \mathbb{R}^n$. Since the norm is continuous,

$$\lim_{j\to\infty} \left\|x^{(k_j)} - a\right\| \leq r \implies \|x - a\| \leq r.$$

Thus $x \in B_r[a]$.

Theorem 3.4.

Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be continuous and $K \subseteq \mathbb{R}^n$ be compact. Then $f(K)$ is also compact.

Proof. Let $(y^{(k)})$ be a sequence in $f(K)$. Therefore there exists a sequence $(x^{(k)})$ in K where $f(x^{(k)}) = y^{(k)}$. Since K is compact, there exists a subsequence (x^{k_j}) that converges to a point $a \in K$. Since f is continuous, then $f(x^{(k_j)}) = y^{(k_j)} \to f(a)$ as $j \to \infty$. Thus the subseqeunce $(y^{(k_j)})$ converges to $f(a) \in f(K)$, hence $f(K)$ is compact. ◇

Def 3.7. Bounded Set

A set $A \subseteq \mathbb{R}^n$ is **bounded** if there exists $M > 0$ such that

$$\|a\| \leq M, \forall a \in A.$$

Theorem 3.5. Compactness $\Leftrightarrow$ Closed and Bounded

Let $K \subseteq \mathbb{R}^n$. Then K is compact iff K is closed and bounded.

Proof.

$\Leftarrow$) Suppose K is closed and bounded. Let $\left(x^{(k)}\right)$ be a sequence of elements in K. Since K is bounded, there exists $M > 0$ such that $\|a\| \leq M$ for all $a \in K$. Therefore $\left\|x^{(k)}\right\| \leq M$ for all $k \geq 0$, thus $\left(x^{(k)}\right)$ is bounded. By Bolzano-Weiestrass, there then exists a subsequence $\left(x^{(k_j)}\right)$ that converges to a point $x \in \mathbb{R}^n$. Since K is closed, $a \in K$. Therefore K is compact. $\diamond$

$\Rightarrow$) Suppose K is compact. Let $a \in \overline{K}$. Then there exists a sequence $\left(x^{(k)}\right)$ of elements in K that converges to a. Since K is compact, there exists a subsequence in K that converges to some $\tilde{a} \in K$. But by the uniqueness of the limit, $a = \tilde{a} \in K$. Therefore $\overline{K} \subseteq K \implies K = \overline{K}$ meaning K is closed. Suppose towards contradiction that K is *not bounded*. Then for any $l \in \mathbb{N}$, there exists $x^{(l)} \in K$ such that $\left\|x^{(l)}\right\| > l$. K is compact therefore there is a subsequence of these terms $\left(x^{(l_j)}\right)$ that converges to some $a \in K$. Since $\left(x^{(k)}\right)$ is convergent, it is bounded. On the other hand, $\left\|x^{(l_j)}\right\| > l_j \geq j$ which means $\left\|x^{(l_j)}\right\| \to \infty$ as $j \to \infty$, a contradiction. Therefore K must be bounded. $\diamond$

Remark. For a general metric space, it is only true in general that K is compact implies K is closed and bounded.

Remark. Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be continuous and $K \subseteq \mathbb{R}^p$ be compact. Then $f^{-1}(K)$ is closed in $\mathbb{R}^n$. However, it need not be compact. For example, consider $f : \mathbb{R} \to \mathbb{R}^2$ where $f(t) = (\cos(t), \sin(t))$. Clearly f is continuous, and $f(\mathbb{R}) = S^1$. However, this means that S^1 which is a compact set under the preimage maps to $\mathbb{R}$, which is not bounded.

Theorem 3.6.

Let $K \subseteq \mathbb{R}^n$ be a compact non-empty set and $f : K \to \mathbb{R}$ be continuous. Then f is bounded and achieves its supremum and infimum. That is $\exists a, b \in K$ such that

$$\sup_{x \in K} f(x) = f(a) \qquad \inf_{x \in K} f(x) = f(b).$$

Proof. Since f is continuous, $f(K)$ is compact and therefore bounded. Hence f is bounded. Note that $f(K) \neq \varnothing$ is bounded. Thus there exists $\sup f(K) = L$. By definition of the supremum, $\forall \varepsilon > 0, \exists x \in K$ such that $L - \varepsilon < f(x) < L$. Take $\varepsilon = \frac{1}{k}$ for $k \in \mathbb{N}$. Then there exists an $x^{(k)}$ for each k such that $L - \frac{1}{k} <$

$f(x^{(k)}) < L$. As $k \to \infty$, it follows $f(x^{(k)}) \to L$. Since $f(x^{(k)})$ is a sequence in $f(K)$ and $f(K)$ is compact and thus closed, $\exists a \in K$ such that $f(a) = L$. A similar argument can be applied to the infimum. $\diamond$

Def 3.8. Uniform Continuity

Let $f : D \to \mathbb{R}^p$ where $D \subseteq \mathbb{R}^n$. Then f is **uniformly continuous** on D if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in D$

$$\|x - y\| \leq \delta \implies \|f(x) - f(y)\| \leq \varepsilon.$$

Example. Consider the distance function $d(x, A)$ for some $A \subseteq \mathbb{R}^n$. Then the function $d(\cdot, A)$ is uniformly continuous. Consider $\varepsilon > 0$ and take $\delta = \varepsilon$. Take $x_0, y \in \mathbb{R}^n$ such that $\|x_0 - y\| \leq \delta = \varepsilon$. Then

$$|d(y, A) - d(x_0, A)| \leq \|y - x_0\| \leq \varepsilon$$

follows from a previous example. Thus the distance function is uniformly continuous.

Theorem 3.7.

Let $K \subseteq \mathbb{R}^n$ be compact and $f : K \to \mathbb{R}^n$ be continuous. Then f is uniformly continuous on K.

Proof. Suppose towards contradiction that f is not uniformly continuous. Then $\exists \varepsilon > 0$ such that $\forall \delta > 0$, there exists $x, y \in K$ where $\|x - y\| < \delta$ while $\|f(x) - f(y)\| > \varepsilon$. Letting $\delta_k = \frac{1}{k}$ for $k \in \mathbb{N}$, there is then corresponding $x^{(k)}$ and $y^{(k)}$ such that $\left\|x^{(k)} - y^{(k)}\right\| \leq \delta_k$ while $\left\|f(x^{(k)}) - f(y^{(k)})\right\|$. By compactness of K, there exists a subsequence $\left(x^{(k_j)}\right)$ that converges to some $x \in K$. Then

$$0 \leq \left\|y^{(k_j)} - x\right\| \leq \underbrace{\left\|y^{(k_j)} - x^{(k_j)}\right\|}_{\leq \frac{1}{k_j} \leq \frac{1}{j}} + \left\|x^{(k_j)} - x\right\|.$$

In the limit as $j \to \infty$, the upper bound goes to 0. Thus $\left\|y^{(k_j)} - x\right\|$ goes to 0, hence $\left(y^{(k_j)}\right)$ converges to x. Since f is continuous at x and y, $f(x^{(k_j)}) \to f(x)$ and $f(y^{(k_j)} \to f(y)$ as $j \to \infty$. Thus

$$\left\|f(x^{(k_j)}) - f(y^{(k_j)})\right\| \leq \left\|f(x^{(k_j)} - f(x)\right\| + \left\|f(x) - f(y^{(k_j)}\right\|$$

which goes to 0 as $j \to \infty$, a contradiction. Thus f is uniformly continuous. $\diamond$

Def 3.9. Open Cover

Let $A \subseteq \mathbb{R}^n$. An **open cover** of A is a collection of open sets (G_α) in $\mathbb{R}^n$ such that $A \subseteq \bigcup G_\alpha$.

Def 3.10. Topological Compactness

A set $K \subseteq \mathbb{R}^n$ is **topologically compact** if every open cover of K has a finite subcover. In other words, for any open cover (G_α) of K, there are $\{\alpha_1, \ldots, \alpha_n\}$ indices with $n < \infty$ such that $K \subseteq G_{\alpha_1} \cup \ldots \cup G_{\alpha_n}$.

Example. The set $I = (0,1) \subseteq \mathbb{R}$ is not topologically compact. Consider the candidate open cover $\bigcup_{x \in (0,1)} \left(\frac{x}{2}, \frac{x+1}{2}\right)$. Let $x \in (0,1)$. Note that

$$\begin{aligned} x > 0 &\implies 2x > x &&\implies x > \frac{x}{2} \\ x < 1 &\implies 2x < x+1 &&\implies x < \frac{x+1}{2} \end{aligned}$$

Thus it is an open cover. Assume then there exists a finite subcover

$$\left(\frac{x_1}{2}, \frac{x_1+1}{2}\right) \cup \ldots \cup \left(\frac{x_n}{2}, \frac{x_n+1}{2}\right)$$

for $x_1, \ldots, x_n \in (0,1)$. Take $x \in \min\{x_1, \ldots, x_n\} > 0$ and $0 < y < \frac{x}{2}$. Then $y \in (0,1)$ but is not in the subcover. Hence I cannot be topologically compact.

3.2.1 Compactness Equivalence

The goal of this section is to prove the following theorem.

Theorem 3.8. Sequential $\Leftrightarrow$ Topological Compactness

A set $K \subseteq \mathbb{R}^n$ is topologically compact iff K is sequentially compact.

The approach will be to use the result being close and bounded is equivalent to sequential compactness as a bridge. That is, show that topological compactness is equivalent to being closed and bounded, and thus sequentially compact as well.

Lemma 3.2.

Let $K \subseteq \mathbb{R}^n$ be (topologically) compact and $F \subseteq K$ be closed in $\mathbb{R}^n$. Then F is also (topologically) compact.

Proof. Let (G_α) be an open cover of F. Note then that $K \subseteq F^c \cup \bigcup_\alpha G_\alpha$. Since F is closed, F^c is open and thus this is an open cover of K. Since K is topologically compact, there then exists $\alpha_1, \ldots, a_n$ finite such that $K \subseteq G_{\alpha_1} \cup \ldots G_{\alpha_n} \cup F^c$. Since $F \subseteq K$, this is a finite cover of F as well. Hence F is compact. ◇

Theorem 3.9. Heine-Borel

Let $K \subseteq \mathbb{R}^n$. Then K is (topologically) compact iff K is closed and bounded.

The following definition and lemma will be pivotal in proving Heine-Borel. If for every compact set K a closed cube Q can be chosen such that $K \subseteq Q$, then by the previous lemma if Q is compact then so is K. Thus the reverse direction of Heine-Borel follows from the compactness of *cubes*.

Def 3.11. Closed Cube

A set $Q \subset \mathbb{R}^n$ is a **closed cube** if there exists closed and bounded intervals $I_1, \ldots, I_n$ in $\mathbb{R}$ such that $Q = I_1 \times \ldots \times I_n$.

Lemma 3.3. Cubes are Compact

Let Q be a closed cube in $\mathbb{R}^n$. Then Q is (topologically) compact.

Lemma 3.4.

Let (I_n) be a sequence of closed bounded intervals in $\mathbb{R}$ such that $I_n \supseteq I_{n+1}$. Then $\bigcap I_n \neq \varnothing$.

Proof. Denote $I_n = [a_n, b_n]$. Note that the set of left endpoints $M = \{a_n : n \in \mathbb{N}\}$ is bounded above by b_1. Let $x = \sup \mathbb{R}$. Note then that

$$a_n \leq a_{n+m} \leq b_{n+m} \leq b_m, \quad \forall n, m \in \mathbb{N}.$$

Thus b_m is an upper bound of M for all $m \geq 1$, meaning $a_m \leq x \leq b_m$ for all $m \geq 1$. Therefore $x \in \bigcap I_n$. ◇

Lemma 3.5.

Let (Q_j) be a sequence of closed cubes in $\mathbb{R}^n$ such that $Q_j \supseteq Q_{j+1}$. Then $\bigcap Q_j \neq \emptyset$.

Proof. Write each Q_j as $I_{1,j} \times \ldots \times I_{n,j}$. Then each $I_{k,j}$ are closed and bounded intervals such that $I_{k,j} \supseteq I_{k+1,j}$. Thus by Lemma 3.4, $\exists y_k \in \mathbb{R}$ for each $1 \leq k \leq n$ such that $y_k \in \bigcap_j I_{k,j}$. Thus the point $y = (y_1, \ldots, y_n) \in \bigcap_j Q_j$. ◇

Proof of Lemma 3.3. Write $Q = [a_1, b_1] \times \ldots \times [a_n \times b_n]$. Suppose towards contradiction that Q is not (topologically) compact. Then there exists an open cover (G_α) of Q that has no finite subcover. Divide Q into 4 subcubes Q_j^1 for $1 \leq j \leq 4$. Since no finite subcover exists for Q, then there is some Q_i^1 that does not have a finite subcovering. Denote $\tilde{Q}_1 = Q$ and $\tilde{Q}_2 = Q_i^1$. The same division and selection process can be applied to $\tilde{Q}_2$ to get some $\tilde{Q}_3$. Continuing gives a sequence $(\tilde{Q}_j)$ such that $\tilde{Q}_j \supseteq \tilde{Q}_{j+1}$, $\tilde{Q}_j$ has no finite subcovering, and

$$\operatorname{diam}(\tilde{Q}_j) := \sup_{x,y \in \tilde{Q}_j} \|x - y\| \leq \frac{\operatorname{diam}(Q)}{2^{j-1}}.$$

for all $j \in \mathbb{N}$. By Lemma 3.5, there is some $y \in \bigcap \tilde{Q}_j$. Since (G_α) is an open cover of Q, there is some G_α with $y \in G_\alpha$. Let $r > 0$ such that $B_r(y) \subseteq Q_\alpha$. Note then if j is taken large enough such that $\frac{\operatorname{diam}(Q)}{2^{j-1}} < r$, then if $x \in \tilde{Q}_j$

$$\|x - y\| \leq \operatorname{diam}(\tilde{Q}_j) \leq \frac{C}{2^{j-1}} < r.$$

Thus $\tilde{Q}_j$ is covered by the single open set G_α, a contradiction. Therefore Q is compact. ◇

Proof of Theorem 3.9.

$\Leftarrow$) Assume K is closed and bounded. Since K is bounded, it is possible to choose a closed cube Q such that $K \subseteq Q$. Since Q is compact by Lemma 3.3 and K is closed, by Lemma 3.2 it follows K is also (topologically) compact.

$\Rightarrow$) Assume K is (topologically) compact. Consider the set of open balls $B_k(0)$ for $k \in \mathbb{N}$ and $k \geq 1$. Note that $\mathbb{R}^n = \bigcup_k B_k(0)$, thus this constitutes an open cover of K. Since K is compact, there must exist a finite subcover

$B_{k_1}(0) \cup \ldots \cup B_{k_m}(0)$. Taking $M = \max\{k_1, \ldots, k_m\}$ gives $K \subseteq B_M(0)$, thus K is bounded.

Consider some $y \in K^c$. Define $G_j = \left\{x \in \mathbb{R}^n : \|x - y\| > \frac{1}{j}\right\}$ for $j \geq 1$. Note that $G_j = \left(B_{j^{-1}}[y]\right)^c$ which means that G_j is open. Since $y \notin K$, then $K \subseteq \mathbb{R}^n \setminus \{y\}$. But $\mathbb{R}^n \setminus \{y\} = \bigcup_{j=1}^{\infty} G_j$ which is then an open cover of K. Since K is compact, there exists a finite subcover $G_{j_1} \cup \ldots G_{j_m}$ of K. Note that taking $j_0 = \max j_1, \ldots, j_m$ gives $G_{j_1} \cup \ldots \cup G_{j_m} \subseteq G_{j_0}$. Therefore $B_{j_0^{-1}}(y) \subseteq G_{j_0}^c \subseteq K^c$. Thus K^c is open meaning K is closed.

$\diamond$

Chapter 4

Differentiability

Def 4.1. Linear Transformation

A map $T : \mathbb{R}^n \to \mathbb{R}^p$ is a **linear transformation** if

$$T(\alpha x + \beta y) = \alpha T x + \beta T y$$

for any $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.

Def 4.2. Standard Basis

The **standard basis** for $\mathbb{R}^n$ is the set of vectors e_k where the i^{th} component is $\delta_{i,j}$. For any $x \in \mathbb{R}^n$, there exists $x_1, \ldots, x_n \in \mathbb{R}$ such that

$$x = x_1 e_1 + \ldots + x_n e_n.$$

Remark. For any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^p$ and $x \in \mathbb{R}^n$, Tx can be written as

$$Tx = T(x_1 e_1 + \ldots + x_n e_n) = x_1 T e_1 + \ldots + x_n T e_n.$$

Denote $Te_k = \left(T_{1k}, \ldots, T_{pk}\right) \in \mathbb{R}^p$. These T_{jk} give a $p \times n$ matrix of the form

$$(T_{ij})_{\substack{j=1,\ldots,p \\ k=1,\ldots,n}} := \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1k} & \cdots & T_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ T_{p1} & T_{p2} & \cdots & T_{pk} & \cdots & T_{pn} \end{bmatrix}.$$

Note then that

$$Te_k = \sum_{j=1}^{p} T_{jk} f_j$$

where f_j is the j^{th} standard basis vector of $\mathbb{R}^p$. Therefore for $x \in \mathbb{R}^n$

$$y = Tx = \sum_{k=1}^{n} x_k \left(\sum_{j=1}^{p} T_{jk} f_j \right) = \sum_{\substack{k=1,\dots,n \\ j=1,\dots,p}} (T_{jk} x_k) f_j.$$

Thus $y_j = T_{j1}x_1 + \ldots + T_{jn}x_n$, meaning

$$\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} T_{11} & \cdots & T_{1k} \\ \vdots & & \vdots \\ T_{p1} & \cdots & T_{pn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Therefore a linear transformation can be identified by this matrix form.

Def 4.3. Linear Transformation Space

Let $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ be the set of all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^p$.

Remark. Recall the definition of differentiability for $f : \mathbb{R} \to \mathbb{R}$ at a point a

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} =: f'(a) \in \mathbb{R}.$$

This definition remains meaningful if $f : \mathbb{R} \to \mathbb{R}^p$, however it does not easily generalize to $f : \mathbb{R}^n \to \mathbb{R}^p$ where $n \geq 2$ since the division would be undefined.

Since the standard definition does not generalize, it is then useful to look at other candidate definitions to generalize. Consider the following alternative characterization for single variable functions.

Theorem 4.1.

Let $f : I \to \mathbb{R}$ where $I = [a, b] \subseteq \mathbb{R}$ and $a \in \mathbb{R}$. Then f is differentiable at a iff $\exists \alpha \in \mathbb{R}$ such that

$$\lim_{h \to 0} \frac{f(x + h) - f(a) + \alpha h}{h} = 0.$$

Moreover $\alpha = f'(a)$.

Proof.

$\Rightarrow$) Suppose f is differentiable at a. Let $\alpha = f'(a)$. Then

$$\lim_{x\to a} \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = 0.$$

Note that if $h = x - a$ that $h \to 0$ as $x \to a$. Thus

$$\lim_{x\to a} \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = \lim_{h\to 0} \frac{f(a+h) - f(a) - \alpha h}{h} = 0.$$

$\Leftarrow$) Suppose $\exists \alpha \in \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - \alpha h}{h} = 0.$$

Let $x = a + h$. Then $x \to a$ as $h \to 0$. Thus

$$0 = \lim_{x\to a} \frac{f(x) - f(a) - \alpha(x-a)}{x-a}$$

$$0 = \lim_{x\to a} \left(\frac{f(x) - f(a)}{x-a} - \alpha\right)$$

$$\alpha = \lim_{x\to a} \frac{f(x) - f(a)}{x-a}$$

Thus f is differentiable at a and $f'(a) = \alpha$.

◇

4.1 Multidimensional Derivative

Def 4.4. Differentiability

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^p$. Then f is **differentiable** at $a \in D$ if there exists $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ such that

$$\lim_{h\to 0} \frac{\|f(a+h) - f(a) - Th\|_{\mathbb{R}^p}}{\|h\|_{\mathbb{R}_n}} = 0$$

where $h \in \mathbb{R}^n \setminus 0$. Furthermore f is differentiable on D if f is differentiable at every $a \in D$.

Theorem 4.2. Uniqueness of Derivative

If $f : D \to \mathbb{R}^p$ is differentiable at a, the corresponding linear transformation $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ is unique.

Proof. Consider any $h \in \mathbb{R}^n \setminus \{0\}$. Replacing h with th where $t \in \mathbb{R}$ and $t \neq 0$ in Def 4.4 gives

$$\rho(t) := \frac{f(a+th) - f(a) - tTh}{|t| \cdot \|h\|}$$

and $\rho(t) \to 0$ as $t \to 0$. Thus

$$\begin{aligned} f(a+th) - f(a) - tTh &= \rho(t)|t|\|h\| \\ Th &= \frac{f(a+th) - f(a)}{t} - \frac{|t|}{t}|h|\rho(t) \end{aligned}$$

Therefore

$$Th = \lim_{t \to 0} \frac{f(a+th) - f(a)}{t} \quad h \in \mathbb{R}^n, h \neq 0$$

meaning T is completely determined by f and is thus unique. ◇

Def 4.5. Differential

If f is differentiable at a, then the **differential** or **total derivative** of f at a is denoted as $f'(a) = \mathcal{D}f(a) = T$.

Example.

1. Consider $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. Note that

$$T(a+h) - T(a) = Ta + Th - Ta = Th.$$

Therefore

$$\lim_{h \to 0} \frac{\|T(a+h) - T(a) - Th\|}{\|h\|} = \lim_{h \to 0} \frac{0}{\|h\|} = 0$$

meaning T is differentiable at every $a \in \mathbb{R}^n$ and is its own differential.

2. Since the identity map $I : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto x$ is a linear transformation, then $I' = I$.

3. Consider $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ where $\rho(x, y) = x \cdot y$. Then ρ is differentiable and $\rho'(a, b)$ is the mapping satisfying

$$\rho'(a, b)(h, k) = a \cdot k + b \cdot h.$$

To see this, note that

$$\begin{aligned}\rho(a+h,b+k)-\rho(a,b) &= (a+h)\cdot(b+k)-a\cdot b\\ &= a\cdot b+h\cdot b+a\cdot k+h\cdot k-a\cdot b\\ &= a\cdot k+b\cdot h+h\cdot k\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\|\rho(a+h,b+k)-\rho(a,b)-\rho'(a,b)(h,k)\|}{\|(h,k)\|} &= \frac{|h\cdot k|}{\|(h,k)\|}\\ &\leq \frac{\|h\|\|k\|}{\|(h,k)\|}\\ &\leq \|k\|\end{aligned}$$

meaning the quantity goes to 0 as $h \to 0$. Thus the proposed differential is indeed correct.

Theorem 4.3.

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. If f is differentiable at $a \in D$, then f is continuous at a.

Theorem 4.4.

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. Then f is differentiable at $a \in D$ iff f_j is differentiable at a_j for $1 \leq j \leq n$.

4.2 Directional and Partial Derivatives

Consider some $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open that is differentiable at $a \in D$. Then for $h \in \mathbb{R}^n$ and $t \in \mathbb{R}$ with $h, t \neq 0$ it follows from Theorem 4.2 that

$$f'(a)h = Th = \lim_{t\to 0}\frac{f(a+th)-f(a)}{t}.$$

Therefore the function $t \mapsto f(a+th)$ is differentiable at 0 with derivative

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(a+th) = f'(a)h.$$

Def 4.6. Directional Derivative

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. Then f has a **directional derivative** at $a \in D$ in the direction $v \in \mathbb{R}^n \setminus \{0\}$ if the function $t \to f(a + tv)$ is differentiable at 0. In this case,

$$f'(a)v = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(a+tv) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t} =: \mathcal{D}_v f(a) =: f'_v(a)$$

is called the directional derivative of f at a in the direction of v.

Def 4.7. Partial Derivative

If $v = e_j \in \mathbb{R}^n$ for $1 \leq j \leq n$, then

$$f'_{e_j}(a) =: \frac{\partial f}{\partial x_j}(a) := \partial_{x_j} f(a)$$

is the j^{th} **partial derivative** of f at a.

Remark. The partial derivative is often taught as holding all but one component constant. This is clear from the definition of partial derivative since

$$\begin{aligned} \frac{\partial f}{\partial x_j} &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(a_1, \ldots, a_{j-1}, a_j + t, a_{j+1}, \ldots, a_n) \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=a_j} f(a_1, \ldots, a_{j-1}, t, a_{j+1}, \ldots, a_n) \end{aligned}$$

Theorem 4.5.

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. Suppose that f is differentiable at $a \in D$. Then

i) f has directional derivatives in all directions $v \in \mathbb{R}^n \setminus \{0\}$ and $f'_v(a) = f'(a)v$.

ii) f has all partial derivatives at a and for all $v \in \mathbb{R}^n$

$$f'(a)v = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}.$$

iii) The matrix form of $f'(a)$ is given by

$$\begin{aligned} f'(a) &= \left[\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right] \\ &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \cdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \cdots & \frac{\partial f_p}{\partial x_n}(a) \end{bmatrix} \end{aligned}$$

Proof.

i) This follows from Def 4.6

ii) Write $v = v_1e_1 + \ldots + v_ne_n$. Since $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ then

$$f'(a)v = \sum_{j=1}^{n} v_j f'(a)e_j = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}.$$

◇

Remark. The converse of (i) does not hold. That is, the existence of all directional derivatives at some point a does not imply differentiability of f at a. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ where

$$f(x_1, x_2) = \begin{cases} \frac{x_1x_2^2}{x_1^2+x_2^4} & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

For any $v \in \mathbb{R}^2$, the directional derivative of f at 0 is

$$\begin{aligned} f_v'(0) &= \lim_{t\to 0} \frac{f(0+tv)-f(0)}{t} \\ &= \lim_{t\to 0} \frac{f(tv)}{t} \\ &= \lim_{t\to 0} \frac{tv_1 t^2 v_2^2}{t(t^2v_1^2 + t^4 v_2^4)} \\ &= \lim_{t\to 0} \frac{v_1 v_2^2}{v_1^2 + t^2 v_2^4} \\ &= \begin{cases} \frac{v_2^2}{v_1} & v_1 \neq 0 \\ 0 & v_1 = 0 \end{cases} \end{aligned}$$

Therefore all directional derivatives exists at 0. However f is not continuous at 0, meaning it cannot be differentiable at 0.

Theorem 4.6.

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. If $\frac{\partial f}{\partial x_j}(x)$ for $1 \leq j \leq n$ exists and is continuous on D, then f is differentiable on D.

Proof. It suffices to consider the case $p = 1$ since differentability of f is equivalent to differentiability of all the component functions $f_j : D \to \mathbb{R}$. Let $x \in D$. The desired result is that

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - \sum_{k=1}^n \frac{\partial f}{\partial x_k}(x) h_k|}{\|h\|} = 0.$$

Note that $f(x+h) - f(x)$ can be written as

$$\begin{aligned} f(x+h) - f(x) = \sum_{k=1}^n \big[& f(x_1+h_1, \ldots, x_k + h_k, x_{k+1}, \ldots, x_n) \\ & - f(x_1 + h_1, \ldots, x_{k-1} + h_{k-1}, x_k, \ldots, x_n) \big] \end{aligned}$$

Let $g(t) = f(\ldots, x_{k-1} + h_{k-1}, t, x_{k+1}, \ldots)$. Then the k^{th} term in the sum is

$$\begin{aligned} & f(x_1 + h_1, \ldots, x_{k-1} + h_{k-1}, x_k + h_k, x_{k+1}, \ldots, x_n) \\ & \quad - f(x_1 + h_1, \ldots, x_{k-1} + h_{k-1}, x_k, x_{k+1}, \ldots, x_n) = g(x_k + h_k) - g(x_k) \end{aligned}$$

Since the partials exists, then g is differentiable and thus by mean value theorem $\exists \psi_k \in (x_k, x_k + h_k)$ such that

$$g(x_k + h_k) - g(x_k) = g'(\psi_k)(x_k + h_k - x_k) = \frac{\partial f}{\partial x_k}(\psi_k) \cdot h_k.$$

Hence

$$\begin{aligned} f(x+h) - f(x) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x) h_k \\ = \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(\ldots, x_{k-1} + h_{k-1}, \psi_k, x_{k+1}, \ldots) - \frac{\partial f}{\partial x_k}(x) \right) h_k \end{aligned} \tag{4.1}$$

Since each partial is continuous at $x \in D$, for any $\varepsilon > 0$ there exists $\delta_k > 0$ such that for any $y \in D$

$$\|x - y\| \leq \delta_k \implies \left\| \frac{\partial f}{\partial x_k}(y) - \frac{\partial f}{\partial x_k}(x) \right\| \leq \varepsilon.$$

Take $\delta = \min\{\delta_1, \ldots, \delta_n\} > 0$. Then for $\|h\| \leq \delta$ since

$$\begin{aligned} \|(\ldots, x_{k-1} + h_{k-1}, \psi_k, x_{k+1}, \ldots) - (x_1, \ldots, x_n)\| &= \|(h_1, \ldots, h_{k-1}, \overbrace{\psi_k - x_k}^{\leq h_k}, 0, \ldots)\| \\ &\leq \|h\| \leq \delta \end{aligned}$$

Thus

$$\left| f(x+h) - f(x) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x) h_k \right| < \varepsilon \sum_{k=1}^{n} |h_k| \leq \varepsilon |h| n.$$

◇

4.3 Chain Rule

Theorem 4.7.

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. Then $\exists c \in \mathbb{R}$ such that for all $x \in \mathbb{R}^n$

$$|Tx|_{\mathbb{R}^p} \leq c|x|_{\mathbb{R}^n}.$$

Proof. Recall that T has a matrix form (T_{jk}) that is $p \times n$. Then by Cauchy-Schwarz

$$|Tx|^2 = \sum_{j=1}^{p}((Tx)_j)^2 = \sum_{j=1}^{p}\left(\sum_{j=1}^{n} T_{jk}x_k\right)^2 \leq \sum_{j=1}^{p}\left(\sum_{k=1}^{n} T_{jk}^2\right)\left(\sum_{k=1}^{n} x_k^2\right) = c^2|x|^2$$

Thus rooting both sides gives $|Tx| \leq c|x|$. ◇

Def 4.8. Continuity Class

Let $D \subseteq \mathbb{R}^n$ be open. Then the **continuity classes** of functions are

- $C^0(D;\mathbb{R}^p) := \{f : D \to \mathbb{R}^p : f \text{ is continuous}\}$
- $C^1(D;\mathbb{R}^p) := \left\{f : D \to \mathbb{R}^p : \frac{\partial f}{\partial x_k} \text{ exists and is continuous on } D, 1 \leq j \leq n\right\}$
- $C^2(D;\mathbb{R}^p) := \left\{f : D \to \mathbb{R}^p : f \in C^1(D;\mathbb{R}^p) \text{ and } \frac{\partial f}{\partial x_j} \in C^1(D;\mathbb{R}^p), 1 \leq k \leq n\right\}$

Theorem 4.8.

$f \in C^1(D;\mathbb{R}^p)$ iff f is differentiable on D and $f' : D \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ is continuous

Proof. ◇

Theorem 4.9. Chain Rule

Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega;\mathbb{R}^p)$, and let $\tilde{\Omega} \subseteq \mathbb{R}^p$ be open such that $f(\Omega) \subseteq \tilde{\Omega}$. Let $g \in C^1(\tilde{\Omega};\mathbb{R}^k)$. Then the composition $h = g \circ f$ is $C^1(\Omega;\mathbb{R}^k)$ and $h'(x) = g'(f(x))f'(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$.

Proof. Since f is differentiable at $x \in \Omega$, then for $k \in \mathbb{R}^n$

$$\lim_{k \to 0} \frac{|f(x+k) - f(x) - f'(x)k|}{|k|} = 0.$$

Let $\rho_1(k) = f(x+k) - f(x) - f'(x)k$. Then f being differentiable at x is equivalent to saying

$$\lim_{k \to 0} \frac{|\rho_1(k)|}{|k|} = 0. \tag{1}$$

Let $l \in \mathbb{R}^p$. Note that since g is differentiable at $f(x) \in \tilde{\Omega}$ that

$$g(f(x)+l) = g(f(x)) + g'(f(x))l + \rho_2(l)$$

where $\lim_{l\to 0} \frac{|\rho_2(l)|}{|l|}$. Choose $l = f(x+k) - f(x)$, and note that $l = f'(x)k + \rho_1(k)$. Then

$$\begin{aligned} h(x+k) &= g(f(x+k)) \\ &= g(f(x)+l) = g(f(x)) + g'(f(x))(f'(x)k + \rho_1(k)) + \rho_2(l) \\ &= h(x) + g'(f(x))f'(x)k + g'(f(x))\rho_1(k) + \rho_2(l) \end{aligned}$$

It follows from (1) that $\exists \delta > 0$ such that $|k| \leq \delta \implies \frac{|\rho_1(k)|}{|k|} \leq 1$. It also follows from Theorem 4.7 that $|f'(x)k| \leq c|k|$ for some $c \in \mathbb{R}$. Therefore $|l| \leq (c+1)|k|$, meaning

$$\begin{aligned} \frac{|g'(f(x))\rho_1(k) + \rho_2(l)|}{|k|} &\leq \frac{|g'(f(x))\rho_1(k)|}{|k|} + \frac{|\rho_2(l)|}{|k|} \\ &\leq c\frac{|\rho_1(k)|}{|k|} + (c+1)\frac{|\rho_2(l)|}{|l|} \end{aligned}$$

which both go to 0 as $k \to 0$ and $l \to 0$. Therefore h is differentiable at $x \in \Omega$ and $h'(x) = g'(f(x))f'(x)$. Since $h'(x)$ is the composition of continuous functions, it follows $h \in C^1(\Omega; \mathbb{R}^k)$. ◇

Example. Suppose $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable. Then the differential of f as a matrix is

$$f'(t) = \begin{bmatrix} \frac{\mathrm{d}f_1}{\mathrm{d}t}(t) \\ \vdots \\ \frac{\mathrm{d}f_n}{\mathrm{d}t}(t) \end{bmatrix}.$$

and the differential of g as a matrix is

$$g'(y) = \begin{bmatrix} \frac{\partial g}{\partial y_1}(y) & \cdots & \frac{\partial g}{\partial y_n}(y) \end{bmatrix} =: \nabla g(y).$$

By the chain rule, the composition $h : \mathbb{R} \to \mathbb{R} : x \mapsto (g \circ f)(x)$ is differentiable and the differential $h' : \mathbb{R} \to \mathcal{L}(\mathbb{R}, \mathbb{R})$ is given by $h'(t) = g'(f(t))f'(t)$. From them matrix forms, it follows

$$h'(t) = \sum_{k=1}^{n} \frac{\partial g}{\partial y_k}(f(t)) \frac{\mathrm{d}f_k}{\mathrm{d}t}(t).$$

Example. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$ and both are differentiable. Then

$$f'(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{bmatrix}$$

and

$$g'(y) = \begin{bmatrix} \frac{\partial g}{\partial y_1}(y) & \cdots & \frac{\partial g}{\partial y_n}(y) \end{bmatrix}.$$

Therefore the composition $h : \mathbb{R}^n \to \mathbb{R} : x \mapsto (g \circ f)(x)$ has differential $h' : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ where

$$\begin{aligned} h'(x) &= \begin{bmatrix} \frac{\partial g}{\partial y_1}(f(x)) & \cdots & \frac{\partial g}{\partial y_n}(f(x)) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^n \frac{\partial g}{\partial y_k}(f(x)) \frac{\partial f_k}{\partial x_1}(x) & \cdots & \sum_{k=1}^n \frac{\partial g}{\partial y_k}(f(x)) \frac{\partial f_k}{\partial x_n}(x) \end{bmatrix} \end{aligned}$$

Def 4.9. Operator Norm

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. Then the **operator norm** of T is

$$\|T\| := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|Tx|_{\mathbb{R}^p}}{|x|_{\mathbb{R}^n}}.$$

Corollary 4.1.

For any $x \in \mathbb{R}^n$, $|Tx| \leq \|T\| |x|$.

Proof. Both (i) and (ii) follow from properties of the Euclidean norm and supremum. Consider then (iii). Note that for $x \in \mathbb{R}^n \setminus \{0\}$

$$|(T+S)x| \leq |Tx| + |Sx| \leq \big(\|T\| + \|S\|\big)|x|.$$

Thus

$$\frac{|(T+S)x|}{|x|} \leq \|T\| + \|S\| \implies \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|(T+S)x|}{|x|} \leq \|T\| + \|S\|$$

which was to be shown. ◇

4.4 Mean Value Theorem

Recall the mean value theorem for single variable real functions.

Theorem 4.10.

Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Then for any $x, x' \in \mathbb{R}$, there exists $\psi \in (x, x')$ such that

$$f(x) - f(x') = f'(\psi) \cdot (x - x').$$

In this current form, the mean value theorem does not generalize to functions $f : \mathbb{R}^n \to \mathbb{R}^p$. Consider $f : \mathbb{R} \to \mathbb{R}^2$ where $f(t) = (\cos t, \sin t)$. Then the derivative is $f'(t) = (-\sin t, \cos t)$. Take $x = 2\pi$ and $x' = 0$. Then

$$f(x) = (1, 0) \qquad f(x') = (1, 0)$$

If the mean value theorem held, then

$$\begin{aligned} f(x) - f(x') &\stackrel{?}{=} f'(\psi)(x - x') \\ (0, 0) &\stackrel{?}{=} 2\pi(-\sin\psi, \cos\psi) \end{aligned}$$

But $|2\pi(-\sin\psi, \cos\psi)| = 2\pi$ and so can never be $(0, 0)$. Thus more care is needed to generalize mean value theorem.

Def 4.10. Convex Set

A set $\Omega \subseteq \mathbb{R}^n$ is **convex** if for all $x, y \in \Omega$ the line segment $tx + (1-t)y \in \Omega$ for all $0 \leq t \leq 1$.

Theorem 4.11. Mean Value Theorem

Let $\Omega \subseteq \mathbb{R}^n$ be open, bounded, and convex. Let $\tilde{\Omega} \subset \mathbb{R}^n$ be open such that $\tilde{\Omega} \supseteq \overline{\Omega}$ and $f \in C^1(\tilde{\Omega}; \mathbb{R}^p)$. Then there exists $M > 0$ such that for all $x, y \in \Omega$ then

$$|f(x) - f(y)| \leq M|x - y|.$$

Proof. Note that $\frac{\partial f}{\partial x_k} \in C(\tilde{\Omega}; \mathbb{R}^p)$. Since $\overline{\Omega}$ is compact, then

$$M := \sup_{x \in \overline{\Omega}} \left| \frac{\partial f}{\partial x_k}(x) \right| < \infty.$$

Recall that the Jacobi matrix of $f'(x)$ is

$$f'(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(x) & \cdots & \frac{\partial f_p}{\partial x_n}(x) \end{bmatrix}$$

and thus for any $x \in \mathbb{R}^n$

$$\|f'\| \leq \left[\sum_{\substack{j=1,\dots,p \\ k=1,\dots,n}} \left(\frac{\partial f_j}{\partial x_k}(x) \right)^2 \right]^{\frac{1}{2}}.$$

Let $x, y \in \Omega$ and $g(t) := f(tx + (1-t)y)$ with $0 \leq t \leq 1$. Since Ω is convex, $g(t)$ is well defined on Ω. Then by the Chain Rule, $g \in C^1([0,1]; \mathbb{R}^p)$ and

$$g'(t) = f'(tx + (1-t)y)(x-y).$$

Since f' is a linear transformation and $(x-y) \in \mathbb{R}^n$

$$|g'(t)| \leq \|f'(tx + (1-t)y)\||x-y| \leq M|x-y|.$$

By the fundamental theorem of calculus

$$|g(1) - g(0)| = \left| \int_0^1 g'(t)\mathrm{d}t \right| \leq \int_0^1 |g'(t)|\mathrm{d}t \leq M|x-y|.$$

Therefore since $g(1) = f(x)$ and $g(0) = f(y)$, meaning

$$|f(x) - f(y)| \leq M|x-y|.$$

$\diamond$

4.5 Scalar Functions

Def 4.11. Scalar Function

A function f is a **scalar function** if it is of the form $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^n$ open.

Def 4.12. Local Extrema

A scalar function $f : D \to \mathbb{R}$ has a **local maximum** at $a \in D$ if there exists an open ball $B_r(a) \subseteq D$ such that for all $x \in B_r(a)$

$$f(x) \leq f(a).$$

Similarly, f has a **local minimum** at $a \in D$ if there exists an open ball $B_r(a) \subseteq D$ such that for all $x \in B_r(a)$

$$f(a) \leq f(x).$$

If f has either a local minimum of maximum at $a \in D$, it is said f has a **local extremum** at a.

Remark. Let $f : D \to \mathbb{R}$ be a scalar function that is differentiable at $a \in D$. Then $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ an the Jacobi matrix of f at a is

$$f'(a) = \begin{bmatrix} \partial_{x_1} f(a) & \cdots & \partial_{x_n} f(a) \end{bmatrix}.$$

This specific case of the differential is called the **gradient** of f and is denoted as $\nabla f(a)$

Theorem 4.12.

Let $f : D \to \mathbb{R}$ be a scalar function differentiable at $a \in D$. Then if f has a local extremum at a then $\nabla f(a) = 0$.

Proof. Suppose f has a local maximum at a. Let $\delta > 0$ such that $R = (a_1 - \delta, a_1 + \delta) \times \ldots \times (a_n - \delta, a_n + \delta) \subseteq D$. Then for any $x \in R$, $f(x) \leq f(a)$. Consider the set of functions

$$g_j : (a_j - \delta, a_j + \delta) \to R : t \mapsto f(\ldots, a_{j-1}, t, a_{j+1}, \ldots).$$

Since f is differentiable at a, then g_j is differentiable at a_j and

$$g_j'(a_j) = \lim_{t \to 0} \frac{g_j(a_j + t) - g_j(a_j)}{t} = \partial_{x_j} f(a_j).$$

Since a is a local maximum, then a_j is a local maximum for g_j. Thus $g'(a_j) = 0$ from single variable analysis. Thus $\partial_{x_j} f(a_j) = 0$ for all $1 \leq j \leq n$, meaning $\nabla f(a) = 0$. A very similar arguement works for a local minimum. $\diamond$

Def 4.13. Critical Point

Let $f : D \to \mathbb{R}$ be a scalar function differentiable at $a \in D$. Then a is a **critical point** for f if $\nabla f(a) = 0$. The associated value $f(a)$ is called the **critical value**.

4.6 Second Order Derivatives

Def 4.14. Second Order Partials

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^n$ open. If the partial derivatives $\partial_{x_k} f$ exist on D, and the partials themselves have partial derivatives, then these derivatives are called the **second order partial derivatives** of f, and are denoted equivalently as

$$\partial^2_{x_i x_k} f \Leftrightarrow \frac{\partial^2 f}{\partial x_i \partial x_k} \Leftrightarrow f''_{x_i x_k}$$

with $1 \leq i, k \leq n$.

Example. The order of partials does matter. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ where

$$f(x) = \begin{cases} x_1 x_2 \left(\frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} \right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

Note that

$$\begin{aligned} \partial_{x_1} f(0, x_2) &= \lim_{t \to 0} \frac{f(0+t, x_2) - f(0, x_2)}{t} \\ &= \lim_{t \to 0} \frac{1}{t} \cdot t \cdot x_2 \left(\frac{t^2 - x_2^2}{t^2 + x_2^2} \right) \\ &= \lim_{t \to 0} x_2 \left(\frac{t^2 - x_2^2}{t^2 + x_2^2} \right) \\ &= -x_2 \end{aligned}$$

Thus $\partial_{x_2}\partial_{x_1} f(0, x_2) = -1$ meaning $\partial^2_{x_2 x_1} f(0,0) = -1$. Now note

$$\begin{aligned}\partial_{x_2} f(x_1, 0) &= \lim_{t\to 0} \frac{f(x_1, 0+t) - f(x_1, 0)}{t} \\ &= \lim_{t\to 0} \frac{1}{t} \cdot t \cdot x_1 \left(\frac{x_1^2 - t^2}{x_1^2 + t^2}\right) \\ &= \lim_{t\to 0} x_1 \left(\frac{x_1^2 - t^2}{x_1^2 + t^2}\right) \\ &= x_1\end{aligned}$$

Thus $\partial_{x_1}\partial_{x_2} f(x_1, 0) = 1$ meaning $\partial^2_{x_1 x_2} f(0,0) = 1$, which does not equal $\partial^2_{x_2 x_1} f(0,0)$.

Theorem 4.13. Equality of Mixed Partials

Let $f : D \to \mathbb{R}^p$ with $D \subseteq \mathbb{R}^p$ open and $a \in D$. Suppose $\partial^2_{x_i x_k} f$ and $\partial^2_{x_k x_i} f$ exist and are continuous in a nbhd of a. Then $\partial^2_{x_i x_k} f(a) = \partial^2_{x_k x_i} f(a)$.

Proof. Using a permutation of indicies, it is suffices to only show that the statement works for $k = 1$ and $i = 2$, and thus also for $n = 2$. Since f is differentiable iff its components are, then it suffices to work with $p = 1$. Thus reconsider f as $f : D \to \mathbb{R}$ with $D \subseteq \mathbb{R}^2$ open and $a \in D$. Let U be a nbhd of a where the first and second order partials exist and are continuous. WLOG, the nbhd U can be described as

$$U = (a_1 - \delta, a_1 + \delta) \times (a_2 - \delta, a_2 + \delta)$$

for some $\delta > 0$. Let $(x_1, x_2) \in U \setminus \{(a_1, a_2)\}$. Let $g : (a_1 - \delta, a_1 + \delta) \to \mathbb{R}$ where $g(t) = f(t, x_2) - f(t, a_2)$. Of interest is the expression

$$[f(x_1, x_2) - f(x_1, a_2)] - [f(a_1, x_2) - f(a_1, a_2)]. \tag{$\star$}$$

Then by MVT, there exists ψ between x_1 and a_1 such that

$$[f(x_1, x_2) - f(x_1, a_2)] - [f(a_1, x_2) - f(a_1, a_2)] = g(x_1) - g(a_1) = g'(\psi)(x_1 - a_1).$$

Again by MVT on $h(t) = \partial_{x_1} f(\psi, t)$, there exists η between x_2 and a_2 such that

$$\begin{aligned}g'(\psi)(x_1 - a_1) &= h'(\eta)(x_1 - a_1)(x_2 - a_2) \\ &= \left[\partial_{x_2}\partial_{x_1} f(\psi, \eta)\right] (x_1 - a_1)(x_2 - a_2)\end{aligned}$$

Consider the following rearrangement of $(\star)$

$$[f(x_1,x_2)-f(x_1,a_2)]-[f(a_1,x_2)-f(a_1,a_2)]=[f(x_1,x_2)-f(a_1,x_2)]-[f(x_1,a_1)-f(a_1,a_2)].$$

By MVT on $\tilde{g}(t)=f(x_1,t)-f(a_1,t)$, there exists $\tilde{\eta}$ between x_2 and a_2 such that

$$\begin{aligned}[f(x_1,x_2)-f(a_1,x_2)]-[f(x_1,a_1)-f(a_1,a_2)] &= \tilde{g}'(\tilde{\eta})(x_2-a_2)\\ &= \left[\partial_{x_2}f(x_1,\tilde{\eta})-\partial_{x_2}f(a_2,\tilde{\eta})\right](x_2-a_2)\end{aligned}$$

Again by MVT on $\tilde{h}(t)=\partial_{x_2}f(t,\tilde{\eta})$, there exists $\tilde{\psi}$ between x_1 and a_1 such that

$$\begin{aligned}\left[\partial_{x_2}f(x_1,\tilde{\eta})-\partial_{x_2}f(a_2,\tilde{\eta})\right](x_2-a_2) &= \tilde{h}'(\tilde{\psi})(x_1-a_1)(x_2-a_2)\\ &= \left[\partial_{x_1}\partial_{x_2}f(\tilde{\psi},\tilde{\eta})\right](x_1-a_1,x_2-a_2)\end{aligned}$$

Therefore since $x_1\neq a_1$ and $x_2\neq a_2$, combining both alternative expressions of $(\star)$ gives

$$\partial^2_{x_2x_1}f(\psi,\eta)=\partial^2_{x_1x_2}f(\tilde{\psi},\tilde{\eta}).$$

The points (ψ,η) and $(\tilde{\psi},\tilde{\eta})$ are in U and since the second partials are continuous at a

$$\lim_{\delta\to 0}\partial^2_{x_2x_1}f(\psi,\eta)=\partial^2_{x_2x_1}f(a_1,a_2)=\partial^2_{x_1x_2}f(a_1,a_2)=\lim_{\delta\to 0}\partial^2_{x_1x_2}f(\tilde{\psi},\tilde{\eta}).$$

$\diamond$

4.7 Taylor's Formula

Theorem 4.14. Taylor's Formula of Second Order

Let $D\subseteq\mathbb{R}^n$ be open with $a\in D$ and $f\in C^2(D;\mathbb{R})$. Then for all $h\in\mathbb{R}^n$ small ($|h|\leq\delta$),

$$f(a+h)=f(a)+L(h)+\frac{1}{2}Q(h)+R(h)$$

where

$$L(h)=\sum_{k=1}^{n}\partial_{x_k}f(a)h_k=\nabla f(a)\cdot h \qquad Q(h)=\sum_{1\leq j,k\leq n}\partial^2_{x_jx_k}f(a)h_jh_k.$$

and

$$\lim_{h\to 0}\frac{R(h)}{|h|^2}=0.$$

Proof. Consider the function $F(t) = f(a + th) \in C^2([0,1];\mathbb{R})$. By the chain rule

$$F'(t) = f'(a + th)h = \nabla f(a + th) = \sum_{k=1}^{n} \partial_{x_k} f(a + th)h_k.$$

Again by the chain rule

$$F''(t) = \sum_{k=1}^{n}\sum_{j=1}^{n} \partial_{x_j}\partial_{x_k} f(a + th)h_j h_k.$$

Therefore $F'(0) = L(h)$ and $F''(0) = Q(h)$. Thus by the one dimensional version of Taylors Formula to F there exists some ψ between 0 and t such that

$$F(t) = F(0) + F'(0)t + \frac{1}{2}F''(\psi)t^2.$$

Plugging in $t = 1$ gives

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(0) = F(0) + F'(0) + F''(0) + \frac{1}{2}F''(0) + R(h)$$

where $R(h) = \frac{1}{2}(F''(\psi) - F''(0))$. Since $F(0) = f(a)$ and $F(1) = f(a + h)$

$$f(a + h) = f(a) + L(h) + \frac{1}{2}Q(h) + R(h).$$

Now consider $R(h)$. Substituting in the expression for $F''(t)$ gives

$$R(h) = \frac{1}{2}\sum_{1\le j,k\le n}\left[\partial_{x_j x_k} f(a + \psi h) + \partial_{x_j x_k} f(a)\right]h_j h_k.$$

Since $\partial_{x_j x_k} f$ is continuous, for any $\varepsilon > 0$ there eists $\delta > 0$ such that for $|x - a| \le \delta$,

$$|\partial_{x_j x_k} f(x) - \partial_{x_j x_k} f(a)| \le \varepsilon.$$

Therefore if $|h| \le \delta$, then $|\partial_{x_j x_k} f(a + \psi h) - \partial_{x_j x_k} f(a)| \le \varepsilon$. Thus this bounds $R(h)$ in that

$$|R(h)| \le \frac{1}{2}\sum_{1\le j,k\le n} \varepsilon |h_j||h_k| \le \frac{1}{2}\sum_{1\le j,k\le n} \varepsilon |h|^2.$$

It follows then that for some $c \in \mathbb{R}$

$$\frac{|R(h)|}{|h|^2} \le c\varepsilon \implies \lim_{h\to 0}\frac{R(h)}{|h|^2} = 0.$$

$\diamond$

Theorem 4.15. Uniqueness of Taylor's Formula

Let $D \subseteq \mathbb{R}^n$ be open with $a \in D$ and $f \in C^2(D;\mathbb{R})$. Suppose for h small that
$$f(a+h) = C + \tilde{L}(h) + \frac{1}{2}\tilde{Q}(h) + \tilde{R}(h)$$
where C is a constant, $\tilde{L}$ is a linear form, $\tilde{Q}$ is a quadratic form, and
$$\lim_{h\to 0} \frac{\tilde{R}(h)}{|h|^2} = 0.$$
Then $C = f(a)$, $\tilde{L} = L$, $\tilde{Q} = Q$ as in from Theorem 4.14

4.8 Quadratic Forms

Let $f \in C^2(D;\mathbb{R})$ for $D \subseteq \mathbb{R}^n$ open. If $a \in D$ is a critical point of f, then $\nabla f(a) = 0$, and by Taylor's Formula of Second Order
$$f(a+h) = f(a) + \frac{1}{2}Q(h) + R(h).$$
Note that this means the behavior of f around small neighborhoods of a is described mostly by $Q(h)$ since $f(a)$ is a constant and $R(h)$ will drop off. Thus it is of interest to understand $Q(h)$ to understand the behavior of f at a.

Def 4.15. Definiteness

Let Q be a quadratic form on $\mathbb{R}^n$. Then Q is

- **positive semidefinite** if $Q(h) \geq 0$ for all $h \in \mathbb{R}^n$
- **positive definite** if $Q(h) > 0$ for all $h \in R^n \setminus \{0\}$
- **negative semidefinite** if $Q(h) \leq 0$ for all $h \in \mathbb{R}^n$
- **negative definite** if $Q(h) < 0$ for all $h \in \mathbb{R}^n \setminus \{0\}$

If Q is neither positive or negative semidefinite, then Q is **indefinite**.

Remark. Q is indefinite iff there exists $h^+, h^- \in \mathbb{R}^n$ such that $Q(h^+) > 0$ and $Q(h^-) < 0$.

An important property of quadratic forms is that they always have an associated matrix form. That is for any quadratic form Q on $\mathbb{R}^n$, there exists a symmetric $n \times n$ matrix A such that

$$Q(h) = Ah \cdot h.$$

The properties of Q can be then ascertained by the properties of its associated matrix A, as outlined below.

Theorem 4.16.

Let Q be a quadratic form. Then for all eigenvalues λ_i of the associated matrix A

- Q is positive definite $\Leftrightarrow \lambda_i > 0$
- Q is positive semidefinite $\Leftrightarrow \lambda_i \geq 0$
- Q is negative definite $\Leftrightarrow \lambda_i < 0$
- Q is negative semidefinite $\Leftrightarrow \lambda_i \leq 0$
- Q is indefinite $\Leftrightarrow$ there exists $\lambda_j > 0$ and $\lambda_k < 0$

4.9 Inverse Functon Theorem

Def 4.16. Contraction

Let $E \subseteq \mathbb{R}^n$. Then a map $f : E \to E$ is a **contraction** if $\exists L \in (0,1)$ such that for all $x, y \in E$

$$|f(x) - f(y)| \leq L|x - y|.$$

Theorem 4.17. Banach's Contraction Mapping Theorem

Let $E \subseteq \mathbb{R}^n$ be closed and let $f : E \to E$ be a contraction. Then f has a unique fixed point $x \in E$ such that $f(x) = x$.

Proof. Let $x^{(0)} \in E$ and define recursively $x^{(k+1)} = f(x^{(k)})$ for $k \geq 0$ [1]. Consider $|x^{(k)} - x^{(m)}|$ for $m > k$. Note that by triangle inequality

$$|x^{(k)} - x^{(m)}| \leq |x^{(k)} - x^{(k+1)}| + |x^{(k+1)} - x^{(k+2)}| + \ldots + |x^{(m-1)} - x^{(m)}|. \quad (\star)$$

Analyzing a single term $|x^{(j)} - x^{(j+1)}|$, since f is a contraction

$$|x^{(j)} - x^{(j+1)}| = |f(x^{(j-1)}) - f(x^{(j)})| \leq L|x^{(j-1)} - x^{(j)}|.$$

Therefore this process can be done inducitvely to get

$$|x^{(j)} - x^{(j-1)}| \leq L^j |x^{(0)} - x^{(1)}|.$$

Substituting this result into $(\star)$ gives

$$\begin{aligned} |x^{(k)} - x^{(m)}| &\leq \sum_{j=k}^{m-1} |x^{(j)} - x^{(j+1)}| \\ &\leq \sum_{j=k}^{m-1} L^j |x^{(0)} - x^{(1)}| \\ &= L^k \cdot |x^{(0)} - x^{(1)}| \cdot \sum_{j=0}^{m-k-1} L^j \\ &\leq \frac{L^k}{1-L} \cdot |x^{(0)} - x^{(1)}| \end{aligned}$$

Since $0 < L < 1$, in the limit as $k \to \infty$, $\frac{L^k}{1-L}$ goes to 0 and thus so does $|x^{(k)} - x^{(m)}|$, hence $\left(x^{(k)}\right)$ is cauchy. Therefore $\left(x^{(k)}\right)$ converges in $\mathbb{R}^n$ to some point $x \in \mathbb{R}^n$. Furthermore, since E is closed and $\left(x^{(k)}\right)$ is a sequence in E it follows $x \in E$. Since f is continuous

$$x^{(k+1)} = f(x^{(k)}) \xrightarrow{k \to \infty} x = f(x).$$

◇

Theorem 4.18. Inverse Function Theorem

Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega; \mathbb{R}^n)$. Let $x_0 \in \Omega$ and $y_0 = f(x_0)$. If $f'(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is invertible, then there exists an open nbhd U of x_0 and an open nbhd V of y_0 such that $f : U \to V$ is a bijection. Furthermore, $f^{-1} : V \to U$ is $C^1(V; \mathbb{R}^n)$.

[1] This technique is called a *Picard iteration*.

Remark. The inverse function theorem only provides a *local bijection*. Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ where $f(x, y) = (e^x \cos y, e^x \sin y)$. The differential of f is

$$f'(x, y) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}.$$

Note that $\det(f'(x, y)) = e^{2x} \neq 0$, therefore the differential is invertible for all $(x, y) \in \mathbb{R}^2$. Therefore by the inverse function theorem every point in $\mathbb{R}^2$ has a nbhd where f is bijective and thus f is injective. But f is not globally injective. In particular

$$f(x, y + 2\pi) = (e^x \cos(y + 2\pi), e^x \sin(y + 2\pi)) = (e^x \cos y, e^x \sin y) = f(x, y).$$

Let g be $f^{-1} : V \to U$. Note by chain rule that

$$g(f(x)) = x \implies g'(f'(x))f'(x) = 1$$

and

$$f(g(y)) = y \implies f'(g(y))g'(y) = 1.$$

Therefore

$$g'(y) = (f'(x))^{-1}.$$

Proof of Theorem 4.18. Assume that $x_0 = y_0 = 0$ and $f'(0) = 1$. Note that $x \to f'(x)$ is continuous at 0 since $f \in C^1$. Therefore there exists $\delta > 0$ such that

$$|x| \leq \delta \implies \|f'(x) - 1\| \leq \frac{1}{2}.$$

For $y \in \mathbb{R}^n$, denote $\phi_y(x) = x + y - f(x)$. Note that $\phi_y \in C^1(\Omega; \mathbb{R}^n)$. Then $y = f(x)$ iff $\phi_y(x) = x$, and $\phi_y'(x) = 1 - f'(x)$, therefore

$$|x| \leq \delta \implies \left\|\phi_y(x)\right\| \leq \frac{1}{2}.$$

By MVT on $|x| \leq \delta$, there exists $M = \sup_{|x| \leq \delta} \left\|\phi_y'(x)\right\| \leq \frac{1}{2}$ such that for all $|x^{(1)}| \leq \delta$ and $|x^{(2)}| \leq \delta$

$$|\phi_y(x^{(1)}) - \phi_y(x^{(2)})| \leq \frac{1}{2}|x^{(1)} - x^{(2)}|. \tag{$\star$}$$

Set $U = \{|x| < \delta\}$ and $V = f(U)$. Since f is injective from U to V, f is bijective. Let $y_0 \in V$ and note that $y_0 = f(x_0)$ for some $x_0 \in U$. Since U is open, then there is an open ball $B = B_r(x_0)$ such that $\overline{B} \subseteq U$. Pick $y \in B_{\frac{r}{2}}(y_0)$. Note that

$$\phi_y(x_0) - x_0 = x_0 + y - f(x_0) - x_0 = y - f(x_0) = y - y_0$$

thus

$$|\phi_y(x_0) - x_0| = |y - y_0| < \frac{r}{2}.$$

Taking $x \in \overline{B}$ gives

$$|\phi_y(x) - x_0| \leq |\phi_y(x) - \phi_y(x_0)| + |\phi_y(x_0) - x_0| \leq \frac{1}{2}|x - x_0| + \frac{r}{2} \leq r.$$

meaning $\phi_y(x) \in \overline{B}$. Therefore by ($\star$), $\phi_y : \overline{B} \to \overline{B}$ is a contraction and thus by Theorem 4.17, there exists a unique fixed point $\tilde{x} \in \overline{B}$. Note then that $\tilde{x} \in U$ and $\phi_y(\tilde{x}) = \tilde{x} \implies y = f(\tilde{x})$. Therefore V is open.

Now consider $g = f^{-1}$. Note that $f'(0) = 1$ is invertible and that $\|f'(x) - 1\| \leq \frac{1}{2}$ for $x \in U$. Let $B \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n)$ such that $\|B - 1\| \leq \frac{1}{2}$. Then for any $v \in \mathbb{R}^n$,

$$|(B-1)v| \leq \|B - 1\||v| \leq \frac{1}{2}|v|.$$

Therefore $|Bv| = |(B-1)v + v \geq |v| - |(B-1)v| \geq \frac{1}{2}|v|$. Thus if $Bv = 0$, then $|Bv| = 0$ meaning $|v| = 0$, hence $v = 0$ and B is one-to-one. Since $N(B) = \{0\}$, by rank nullity theorem

$$\text{nullity}(B) + \text{rank}(B) = n \implies \text{rank}(B) = n$$

thus B is onto and hence is a bijection. Therefore $\|f'(x) - 1\| \leq \frac{1}{2}$ for all $x \in U$, then $(f'(x))^{-1}$ exists. Take $y = f(x) \in V$ and let $k \in \mathbb{R}^n$ be small such that $y + k \in V$. Then $g(y) = f^{-1}(f(x)) = x$ meaning $y + k = f(\tilde{x})$ where $\tilde{x} = x + h$ for some $h \in \mathbb{R}^n$. Therefore $g(y+k) = x + h$. Consider $g(y+k) - g(y) - Ak$ where $A = (f'(x))^{-1}$. Note that

$$\begin{aligned}
g(y+k) - g(y) - Ak &= x + h - x - Ak \\
&= h - Ak \\
&= -A\Big(k - A^{-1}h\Big) \\
&= -A(f(x+h) - f(x) - f'(x)h)
\end{aligned}$$

Therefore

$$\frac{g(y+k)-g(y)-Ak}{|k|} = \frac{-A(f(x+h)-f(x)-f'(x)h)}{|h|} \cdot \frac{|h|}{|k|}.$$

Note that

$$\begin{aligned}\phi_y(x+h)-\phi_y(x) &= y+x+h-f(x+h)-(y+x-f(x)) \\ &= h+f(x)-f(x+h) \\ &= h-k\end{aligned}$$

By ($\star$), it follows that

$$|h|-|k| \le |h-k| = |\phi_y(x+h)-\phi_y(x)| \le \frac{1}{2}|x+h-x| = \frac{1}{2}|h|$$

meaning $\frac{1}{2}|h| \le |k|$. Therefore as $k \to 0$, then $h \to 0$ and $\frac{|h|}{|k|} \le 2$. In total then

$$\lim_{k\to 0} \frac{g(y+k)-g(y)-Ak}{|k|} = \lim_{k\to 0} \frac{-A(f(x+h)-f(x)-f'(x)h)}{|h|} \cdot \frac{|h|}{|k|} = 0$$

hence g is differentiable. By Cramer's rule,

$$g'(y) = \frac{1}{\det(f'(g(y)))} \cdot C$$

where C is the cofactor matrix of $f'(g(y))$, which will be filled with continuous entries. Therefore $g \in C^1(V;U)$. ◇

Corollary 4.2.

Let $f \in C^1(\Omega, \mathbb{R}^n)$ with $\Omega \subseteq \mathbb{R}^n$ open be such that $\det f'(x) \neq 0$ for all $x \in \Omega$. Then f is an **open map**. That is, for any open set $\omega \subseteq \Omega$, the set $f(\omega)$ is open in $\mathbb{R}^n$.

Proof. Let $\omega \subseteq \Omega$ be open and $y_0 \in f(\omega)$. Then $y_0 = f(x_0)$ for some $x_0 \in \omega$. Since then $\det f'(x_0) \neq 0$, by Theorem 4.18 there exists open nbhds $U \subseteq \omega$ of x_0 and V of y_0 such that $f : U \to V$ is bjiective. Therefore $y_0 = f(x_0) \in f(U) = V \subseteq f(\omega)$. Therefore y_0 is in an open set, namely V, which is contained in $f(\omega)$. Thus $f(\omega)$ is open. ◇

Corollary 4.3.

Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega, \mathbb{R}^n)$. If $\det f'(x) \neq 0$ for all $x \in \Omega$ and f is injective, then $f(\omega)$ is open and $f : \Omega \to f(\Omega)$ is in C^1 and is bijective with a C^1 inverse.

4.10 Implicit Function Theorem

Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$ where $f(x_1, x_2) = x_1^2 + x_2^2 + 1$. Note that the set $S = \left\{x \in \mathbb{R}^2 : f(x) = 0\right\}$ is the same as

$$S = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\right\}$$

which geometrically is the unit circle. Let $a \in S$ where $a_2 \neq 0$. Therefore for all x in a nbhd of a,

$$f(x_1, x_2) = 0 \Leftrightarrow x_2 = \Psi(x_1), \qquad \Psi(x_1) = \begin{cases} \sqrt{1 - x_1^2} & a_2 > 0 \\ -\sqrt{1 - x_1^2} & a_2 < 0 \end{cases}.$$

However if $a_2 = 0$, then there does not exist a nbhd of a such that $f(x_1, x_2) = 0$ can be given as the graph of a function of x_1. Note that

$$\partial_{x_2} f(a_1, a_2) = 2a_2 = 0 \Leftrightarrow a_2 = 0.$$

That is, the points that posed problems are those where the derivative of f vanishes with respect to x_2.

Theorem 4.19. Implicit Function Theorem

Let $\Omega \subseteq \mathbb{R}^{n+m} = \mathbb{R}^n_x \times \mathbb{R}^m_y$ be open and $f \in C^1(\Omega, \mathbb{R}^n)$. Let $(x_0, y_0) \in \Omega$ be such that $f(x_0, y_0) = 0$. If $f'_x(x_0, y_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is invertible, then

- There exists a nbhd U of x_0 in $\mathbb{R}^n$ and a nbhd V of y_0 in $\mathbb{R}^m$ such that for every $y \in V$, there exists a unique $x \in U$ where $f(x, y) = 0$
- There exists a map $\Psi : V \to U$ in C^1 where $\Psi(y) = x$ and $f(\Psi(y), y) = 0$
- $\Psi'(y) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ is given by $\Psi'(y) = -\left(f'_x(\Psi(y), y)^{-1}\right) \circ f'_y(\Psi(y), y)$ for all $y \in V$

Example. Consider $f(x,y) = x^3y^2 - 3xy^3 - 2x^2$. Note that $f(3,1) = 27 - 9 - 18 = 0$. The differential of f with respect to y is $f_y'(x,y) = 2x^3y - 9xy^2$ and $f_y'(3,1) = 27 \neq 0$. Therefore implicit function theorem applies, meaning the equation $f(x,y) = 0$ defines y as $y = y(x) \in C^1$ in a nbhd of 3. Note that $f(x,y(x)) = 0$, meaning for any x in a nbhd of 1

$$x^3y(x)^2 - 3xy(x)^3 - 2x^2 = 0.$$

Differentiating this with respect to x gives

$$3x^2y(x)^2 + 2x^3y(x)y'(x) - 3y(x)^3 - 9xy(x)^2y'(x) - 4x = 0.$$

Note then that plugging in $x = 3$ and $y(3) = 1$ gives

$$27 + 54y'(3) - 3 - 27y'(3) - 12 = 0 \implies y'(3) = -\frac{4}{9}.$$

Proof of Theorem 4.19. `TODO: Finish, this is only a sketch of the proof`

Define $F : \Omega \to \mathbb{R}^n \times \mathbb{R}^m$ where $F(x,y) = (f(x,y), y)$. Then $F(x_0,y_0) = (0,y_0)$ and

$$F'(x,y) = \begin{bmatrix} f_x'(x,y) & f_y'(x,y) \\ 0 & 1 \end{bmatrix}.$$

Since $f_x'(x_0,y_0)$ is invertible, then so is $F'(x_0,y_0)$ since $F'(x,y)$ is a diagonal matrix and thus its determinant is $f_x'(x,y) \cdot 1 = f_x'(x,y)$ which is invertible. By IVT, there is a nbhd $W_0 \subseteq \Omega$ of (x_0,y_0) and a nbhd V_0 of $(0,y_0)$ such that $F : W_0 \to V_0$ is bijective with a C^1 inverse $G : V_0 \to W_0$. For any (z,y), it is possible to write

$$(z,y) = F(G(z,y)) = F(G_1(z,y), G_2(z,y)) = (f(G_1(z,y), G_2(z,y)), G_2(z,y)).$$

Thus $y = G_2(z,y) \implies G(z,y) = (G_1(z,y), y)$, meaning $z = f(G_1(z,y), y)$. Therefore since z can be any point in a nbhd of 0, taking $z = 0$ gives $f(G_1(0,y), y) = 0$, giving $\Psi(y) := G_1(0,y)$. $\diamond$